

Cite this: *Chem. Sci.*, 2018, **9**, 2740

L718Q mutant EGFR escapes covalent inhibition by stabilizing a non-reactive conformation of the lung cancer drug osimertinib†

D. Callegari, ^a K. E. Ranaghan, ^b C. J. Woods, ^b R. Minari, ^c M. Tiseo, ^c M. Mor, ^a A. J. Mulholland^b and A. Lodola ^{*a}

Osimertinib is a third-generation inhibitor approved for the treatment of non-small cell lung cancer. It overcomes resistance to first-generation inhibitors by incorporating an acrylamide group which alkylates Cys797 of EGFR T790M. The mutation of a residue in the P-loop (L718Q) was shown to cause resistance to osimertinib, but the molecular mechanism of this process is unknown. Here, we investigated the inhibitory process for EGFR T790M (susceptible to osimertinib) and EGFR T790M/L718Q (resistant to osimertinib), by modelling the chemical step (*i.e.*, alkylation of Cys797) using QM/MM simulations and the recognition step by MD simulations coupled with free-energy calculations. The calculations indicate that L718Q has a negligible impact on both the activation energy for Cys797 alkylation and the free-energy of binding for the formation of the non-covalent complex. The results show that Gln718 affects the conformational space of the EGFR–osimertinib complex, stabilizing a conformation of acrylamide which prevents reaction with Cys797.

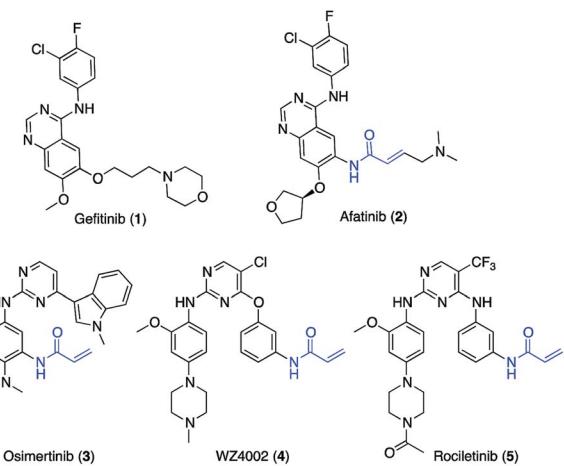
Received 3rd November 2017
Accepted 3rd February 2018

DOI: 10.1039/c7sc04761d
rsc.li/chemical-science

Introduction

Epidermal growth factor receptor (EGFR) is a transmembrane protein which possesses an extracellular EGF binding domain and an intracellular tyrosine kinase domain.¹ EGFR activation by its physiological ligand EGF leads to receptor dimerization and phosphorylation, two events that trigger the activation of signal transduction cascades promoting cell proliferation.² In non-small cell lung cancer (NSCLC), overexpression of EGFR or hyper-activating mutations in its kinase domain have been observed in at least 50% of cases.³ More than 90% of the known kinase mutations occur as short in-frame deletions in exon 19 or as point mutations in exon 21, the latter resulting in arginine replacing leucine at codon 858 (L858R).⁴ These mutations are critical for NSCLC insurgence and progression as their presence results in constitutive activation of EGFR regardless of the presence of EGF.⁵ First-generation EGFR inhibitors (*e.g.*, gefitinib, **1**, Scheme 1) are widely employed as first-line therapy for NSCLC with activating EGFR mutations (*i.e.* L858R

substitutions in exon 21).⁶ Although patients show good responses to the therapy, most of them acquire drug resistance within 1 year treatment, which is driven in about 60% of cases by an additional EGFR T790M point mutation⁷ occurring at the gatekeeper position of the ATP binding site. The change in the steric and lipophilic property of the gatekeeper residue is likely to be responsible for the reduced inhibitory potency of first-generation EGFR inhibitors.^{8,9} The second generation of EGFR inhibitors, such as afatinib (**2**, Scheme 1), demonstrated promising activity against T790M-positive tumors in preclinical models.¹⁰ Due to the presence of an acrylamide warhead



^aDepartment of Food and Drug, University of Parma, Parma, Italy. E-mail: alessio.lodola@unipr.it

^bSchool of Chemistry, University of Bristol, Bristol, UK

^cMedical Oncology Unit, University Hospital of Parma, Parma, Italy

† Electronic supplementary information (ESI) available: p*K*_a shift for Cys797; geometries of TSs identified with QM/MM calculations; analysis of the minimum free-energy path for Cys797 alkylation; analysis of MD replicas; convergence for US simulations; replica of simulation of Cys797 alkylation; conformational FESs obtained from each MD replica. See DOI: 10.1039/c7sc04761d

Scheme 1 Structures of relevant EGFR inhibitors.



capable of alkylating Cys797,¹¹ afatinib can circumvent ATP competition and thus overcome the unfavorable effect caused by the presence of methionine at the gatekeeper position.¹¹ Nevertheless, in the clinic afatinib showed dose-limiting toxicity resulting from potent inhibition of the wild-type (wt) form of EGFR.¹²

Osimertinib (3, Scheme 1)¹³ is a third-generation EGFR inhibitor approved for patients affected by metastatic EGFR T790M mutation-positive NSCLC, who have progressed on or after the therapy with first- and second-generation EGFR inhibitors.¹⁴

Like other third-generation inhibitors, such as WZ4002 (4) and rociletinib (5), osimertinib possesses a 2-aminopyrimidine scaffold which confers selectivity for the oncogenic forms of EGFR *versus* the wt, and an acrylamide group that alkylates Cys797 ensuring the ability to potently inhibit EGFR also in the presence of the T790M mutation.¹⁵

The development of novel forms of resistance is currently limiting the clinical therapeutic benefit of osimertinib.¹⁶ The C797S mutation, which replaces the cysteine with the less nucleophilic serine, has emerged as the main determinant of resistance to third generation EGFR inhibitors.¹⁷ Other mutations, *i.e.*, L718Q occurring in the P-loop and L844V, occurring in the ATP binding site, were initially reported to confer resistance to pyrimidine-based inhibitors, including WZ4002 and rociletinib, *in vitro*.¹⁸ Moderate activity on cells harboring these mutations was maintained by osimertinib, suggesting that it might be still clinically effective in patients harboring L718Q or L844V mutation. However, a cell proliferation assay conducted in Ba/F3 cells expressing L858R/T790M/L718Q EGFR mutant, showed that osimertinib was considerably less potent (~100-fold) at inhibiting cell growth than in Ba/F3 cells expressing either L858R/T790M double mutant or L858R/T790M/L844V triple mutant. This suggests that L718Q mutation likely affects the ability of osimertinib to irreversibly inhibit EGFR. As a matter of fact, clinical resistance to osimertinib has recently been reported in a NSCLC patient expressing EGFR L858R/T790M who also acquired the L718Q mutation.¹⁹ L718Q thus emerges as a mutation able to reduce osimertinib potency *in vitro* and to confer NSCLC resistance *in vivo*.¹⁹ Starting from the visual inspection of the X-ray structure of EGFR–osimertinib complex,²⁰ it has been proposed that the replacement of a leucine with a glutamine at the position 718 of EGFR could reduce the affinity of EGFR for osimertinib and/or hinder Cys797 alkylation by the acrylamide warhead.¹⁹ Considering that Gln718 occupies a peripheral position of the ATP binding site of EGFR (Fig. 1), the precise role of Gln718 in reducing osimertinib activity remains largely unexplained.

Computational simulation protocols able to rationalize the effects of a mutation in the binding site of EGFR have the potential to contribute significantly to anti-cancer drug discovery;²¹ such methods should help in the design of novel inhibitors able to circumvent mutations. Given the increasing interest in and clinical importance of covalent inhibitors, there is a particular need for methods able to analyze effects of mutation on covalent inhibition.

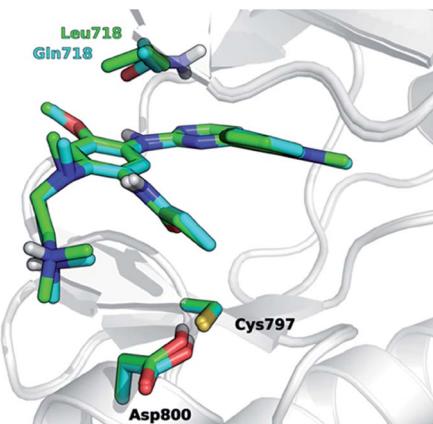


Fig. 1 Superposition of 3D models, derived from the X-ray structure 4ZAU.pdb,²⁰ representing osimertinib in non-covalent complexes with EGFR T790M (green carbon atoms) and EGFR T790M/L718Q mutants (cyan carbon atoms), respectively.

In the present work, we investigate the impact of L718Q mutation on the inhibitory activity of osimertinib by modelling both its chemical step (*i.e.*, Cys797 alkylation) and recognition step (*i.e.* formation of the non-covalent, pre-reactive EGFR–osimertinib complex). We started from X-ray²⁰ derived models of osimertinib in complex with EGFR T790M and EGFR T790M/L718Q mutants. We assessed the influence of the L718Q mutation in EGFR inhibition, estimating for the two molecular systems: (i) the preferred ionization state for Cys797; (ii) the energetics for Cys797 alkylation; (iii) the free-energy of binding for the formation of the non-covalent complex; (iv) the conformational space explored by osimertinib within the two considered EGFR variants. An array of diverse and complementary computational methods was applied. These include molecular dynamics (MD) simulations,²² coupled with umbrella sampling (US),²³ hybrid quantum mechanics molecular mechanics (QM/MM)²⁴ and replica-exchange/thermodynamic integration (RETI)²⁵ approaches.

Results and discussion

Ionization state of Cys797

The reactivity of cysteines with electrophilic compounds depends first of all on the protonation state of their thiol group, as indicated by *e.g.*, the pH dependence of the reaction rate for covalent bond formation.²⁶ EGFR has a solvent-exposed cysteine (Cys797) that is the site of covalent modification by acrylamide-containing irreversible EGFR inhibitors.^{6,27} The Cys797 thiol is not only readily alkylated by Michael acceptors, but is also easily transformed into sulfenic acid (*i.e.*, Cys-S-OH) in the presence of oxidative stimuli.²⁷ This evidence to date supports the hypothesis that the Cys797 thiol exists in the anionic form at physiological pH. This could be due to the presence of an organized microenvironment²⁸ in EGFR able to stabilize its negative charge.²⁹ Furthermore, the Cys797 side chain makes contacts with Asp800, with the sulfur atom at a short distance from one of the oxygen of the carboxylic group, *i.e.*, 3.3 Å in the EGFR–



osimertinib structure.²⁰ Thus, the carboxylate function of Asp800 is probably the natural acceptor of the Cys-SH proton (Scheme 2).

Previous free-energy simulations have shown that, in the presence of a *N*-(4-anilinoquinazolin-6-yl)acrylamide inhibitor, Cys797-S[−]/Asp800-COOH pair is significantly more stable than Cys797-SH/Asp800-COO[−].³⁰ Moreover, titration experiments on wt-EGFR showed that Cys797 has a *pK_a* of 5.5,³¹ a value significantly lower than that of a free thiol in solution (≈ 8.0), and approaching that of a carboxylic acid. This supports the computational finding that the most abundant protonation state for this cysteine/aspartate pair of EGFR is Cys797-S[−]/Asp800-COOH. Therefore, this was the protonation state employed in our models; we also investigated *pK_a* shifts of Cys797 and modeled the proton transfer between it and Asp800.

We began by evaluating the impact of the L718Q mutation on the *pK_a* of Cys797. With this aim, we built molecular models of EGFR T790M and T790M/L718Q mutants using the X-ray structure of osimertinib in non-covalent complex with wt-EGFR.²⁰ The models were equilibrated by molecular dynamics (MD) simulations (see Experimental section for details). For multiple snapshots taken from the MD simulations, *pK_a* calculations performed with PropKa (which uses an empirical scoring function)³² and H++ (which solves the Poisson–Boltzmann equation)³³ indicated that L718Q mutation has no significant effect on the acidity of Cys797 thiol group, with negligible *pK_a* shifts calculated by both methods (0.02 ± 0.38 for PropKa and 0.47 ± 0.43 for H++, respectively; see also Table S1, ESI†). It should be pointed out that these ΔpK_a values are simple empirical estimates and thus they may suffer from limited accuracy.

We next investigated if the L718Q mutation could affect the prototropic equilibrium of the Cys797/Asp800 pair. Applying a SCC-DFTB/AMBER99SB QM/MM potential³⁴ with umbrella sampling (US) MD simulations,²³ we calculated the free-energy difference between the Cys797-S[−]/Asp800-COOH and Cys797-SH/Asp800-COO[−] states. Starting from the molecular models of osimertinib in complex with EGFR T790M or the EGFR T790M/L718Q variant, the proton present on the carboxylic group of Asp800 was moved toward the thiolate group of Cys797 in US MD (see Experimental section).

SCC-DFTB/AMBER99SB QM/MM US MD simulations showed that the Cys797-S[−]/Asp800-COOH state was of similar stability to the Cys797-SH/Asp800-COO[−] state for EGFR T790M (estimated difference of 0.5 ± 0.2 kcal mol^{−1}), while in the case of EGFR T790M/L718Q, the Cys797-S[−]/Asp800-COOH state was more stable by 2.0 ± 0.3 kcal mol^{−1}. It appears that the L718Q mutation actually stabilizes the thiolate form of Cys797. It is apparent that the effects of this mutation on inhibition by osimertinib alkylation are not caused by an increase in the *pK_a* of Cys797.

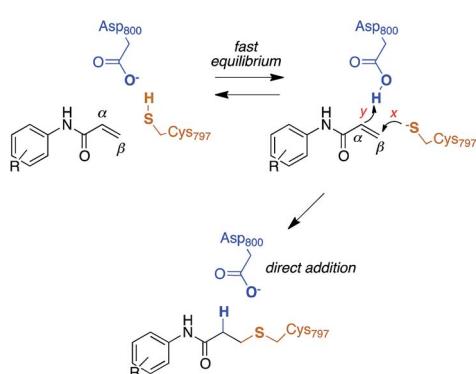
Reaction energetics for Cys797 alkylation

Computational studies recently performed by us³⁰ and others³⁵ indicate that acrylamide-based inhibitors of EGFR alkylate Cys797 by a direct addition mechanism in which the thiolate group of Cys797 attacks the β carbon of the warhead, while the carboxylic acid group of Asp800 protonates the α carbon, leading to a stable 3-(alkylsulfanyl)propanamide adduct (Scheme 2). Using this reaction scheme, we modelled the Michael-type addition of Cys797 to osimertinib, for both EGFR T790M and EGFR T790M/L718Q mutants.

To analyse the reaction energetics, the free-energy surface (FES) of the direct addition mechanism was calculated from US simulations at the SCC-DFTB/AMBER99SB level, using simple reaction coordinates calculated from distances between the reactive atoms: $S_{\text{Cys797}-\text{C}\beta_{\text{acrylamide}}}$ distance modeled the nucleophilic attack (event **x**), and the difference of distances $[(\text{H}_{\text{Asp800}}-\text{C}\alpha_{\text{acrylamide}}) - (\text{H}_{\text{Asp800}}-\text{O}_{\text{Asp800}})]$ described the protonation at $\text{C}\alpha$ of acrylamide (event **y**). While the SCC-DFTB method is known to underestimate absolute barriers in many cases, this approach has been satisfactorily applied to elucidate the effects of mutations on several enzyme-catalyzed reactions; it provides useful and predictive relative barriers.³⁶

The FES is a two-dimensional projection of a chemical process that involves simultaneous changes in two coordinates. Analysis of the FES is useful to explain how changes in the microenvironment of the enzyme affect the position and geometry of the transition state (TS), and thus the energy barrier of the investigated process.³⁷ The SCC-DFTB/AMBER99SB FES for the alkylation of Cys797 by osimertinib for EGFR T790M is shown in Fig. 2A, and that for EGFR T790M/L718Q in Fig. 2B. In both cases, the minimum free-energy path connecting reactants (**R**, upper right corner) and the products (**P**, bottom left corner) follows a diagonal pathway, indicating that nucleophilic attack and protonation of the incipient carbanion species are tightly coupled events. No carbanion/enolate species was identified as a stable minimum, consistent with previous DFT calculations on the addition of thiolates to acrylamides.^{30,35}

The activation free-energy (ΔA_{act}) at the SCC-DFTB/AMBER99SB level of theory is 7.8 ± 0.2 kcal mol^{−1} for alkylation of EGFR T790M and of 8.1 ± 0.2 kcal mol^{−1} for alkylation of EGFR T790M/L718Q. The barrier to the forward reaction (formation of the covalent complex) is unaffected by the L718Q mutation. The reaction free-energy (ΔA_{reac}) is highly negative: -10.3 ± 0.2 and -12.2 ± 0.2 kcal mol^{−1} for EGFR T790M and EGFR T790M/L718Q, respectively, indicating that Cys797



Scheme 2 Reaction mechanism of Cys797 alkylation for acrylamide inhibitors of EGFR.



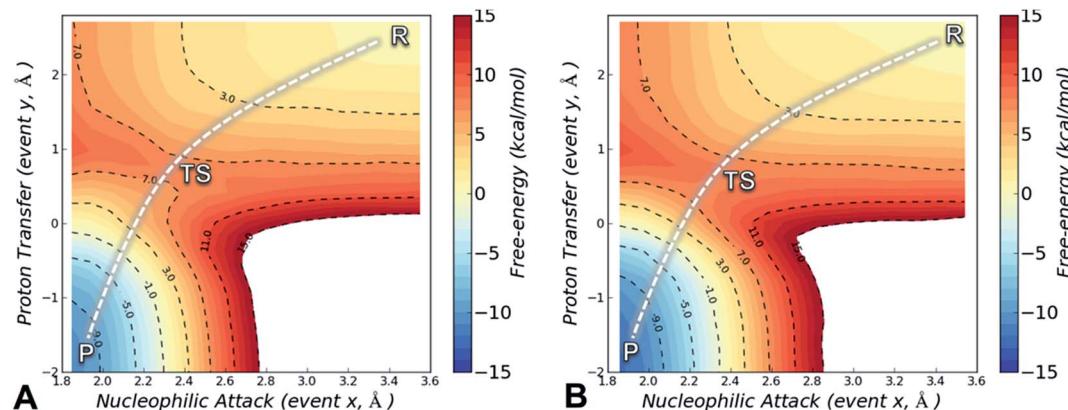


Fig. 2 Free-energy surfaces for Cys797 alkylation by osimertinib in EGFR T790M (panel A) and in EGFR T790M/L718Q (panel B), calculated from umbrella sampling simulations at the SCC-DFTB/AMBER99SB level of QM/MM theory. The reaction coordinates (nucleophilic attack and proton transfer), are given in Å. Free energies are given in kcal mol⁻¹, and the contour levels are set at 1 kcal mol⁻¹ while dashed-contour lines are set every 4 kcal mol⁻¹.

alkylation by osimertinib is highly exergonic and thus a spontaneous process. The covalent adduct (products) is predicted to be slightly more stable in the L718Q mutant, thus the barrier to the reverse reaction is a little higher, but this is not expected to be significant as covalent inhibition is likely to be effectively irreversible. Overall, these results indicate that L718Q mutation does not significantly change the energetics of the alkylation. Consistent with this finding, analysis of the geometries identified along the minimum free-energy path connecting **R** to **P** indicates that L718Q mutation has a negligible impact on the TS geometries obtained by US simulation (Table S2†).

The conformation adopted by the two conjugated double bonds of the acrylamide fragment, which remained *s-cis* during the entire alkylation process, is similar in the two reaction paths. Further analysis of the minimum free-energy paths (Fig. S1 and S2†), revealed that for both systems the key event of the reaction was the nucleophilic attack of the Cys797 sulfur atom on the acrylamide Cβ, which required complete desolvation of the thiolate anion. In agreement with this finding, structures of the TS for Cys797 alkylation of EGFR T790M, show that the formation of S-Cβ bond is quite advanced, with an average (S-Cβ) distance of 2.41 ± 0.05 Å. In contrast,

protonation of the Cα by Asp800 was not very advanced at the TS, with an average H-Cα distance of 1.85 ± 0.08 Å (Fig. 3A). In the case of the EGFR T790M/L718Q variant, the TS structures were slightly more advanced toward the products, as indicated by S-Cβ and H-Cα distances of 2.25 ± 0.07 Å and 1.48 ± 0.04 Å, respectively (Fig. 3B). These minor differences in the average geometries of the TS for the direct addition mechanism explain the negligible difference (0.3 kcal mol⁻¹) in the computed activation free-energies ΔA_{act} . Similar geometries were also found for the products of the reaction.

The thioether adduct **P** was characterized by S-Cβ and H-Cα distances of 1.83 ± 0.03 Å and 1.11 ± 0.04 Å, respectively for EGFR T790M. The corresponding S-Cβ and H-Cα distances were 1.84 ± 0.02 Å and 1.10 ± 0.03 Å for the T790M/L718Q variant, respectively. Thus, analysis of reaction paths shows no significant difference for the two EGFR mutants. All the simulation evidence thus indicates that the L718Q mutation does not reduce the reactivity of Cys797. It is important to note that the starting point (reactant/non-covalent complex) for the QM/MM calculations is a reactive conformation in which the nucleophilic sulfur is in close proximity to the electrophile; these simulations therefore do not address possible changes in favored conformations in the non-covalent complex, which are investigated below.

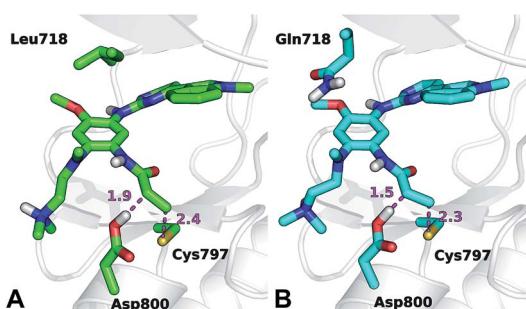


Fig. 3 Representative structures of the TS for Cys797 alkylation in the case of EGFR T790M (panel A, green carbon atoms) and EGFR T790M/L718Q (panel B, cyan carbon atoms). Important distances (in Å) are shown by dashed magenta lines.

Binding affinity for the formation of osimertinib/EGFR mutant non-covalent complexes

It has been recently proposed that the main effect of the replacement of Leu718 with Gln718 could be to disrupt the beneficial hydrophobic/steric interactions involving the methoxyphenyl moiety³⁸ of third generation EGFR inhibitors (Fig. 1) and thus to hamper the formation of the non-covalent complex through a significant reduction of inhibitor affinity. We tested this hypothesis *in silico* by calculating the binding free-energy (ΔA_{bind}) of osimertinib for EGFR T790M and EGFR T790M/L718Q, using the WaterSwap method³⁹ in combination with conventional MM MD simulations (see Experimental



section) performed using the AMBER99SB force field.⁴⁰ WaterSwap estimates ΔA_{bind} using a replica-exchange thermodynamic integration (RETI) algorithm along a reaction coordinate that switches the ligand of interest with an equivalent volume of water in the protein binding site.³⁹ While this methodology tends to overestimate the values of ΔA_{bind} , it is able to capture the effect of mutations occurring in the binding site or in its surroundings.⁴¹ To obtain a robust estimate of the binding free-energy, ten independent WaterSwap calculations were performed for each system *i.e.*, starting from ten different protein-ligand structures taken from a 300 ns-long MD simulation. WaterSwap simulations indicate (Table S3†) that osimertinib has similar ΔA_{bind} for EGFR T790M ($-34.9 \pm 0.7 \text{ kcal mol}^{-1}$) and EGFR T790M/L718Q ($-34.8 \pm 1.9 \text{ kcal mol}^{-1}$) suggesting that L718Q mutation does not significantly affect inhibitor affinity.

Dynamics of osimertinib/EGFR mutant complexes

We finally investigated whether L718Q mutation affects the conformational space of osimertinib in the EGFR mutant- osimertinib complexes. We performed extensive MD simulations (4 independent replicas of 300 ns each) of osimertinib in non-covalent complex with EGFR T790M and with EGFR T790M/L718Q, using the AMBER99SB force field⁴⁰ (see Experimental section for details). We then analyzed the simulations to assess the influence of the mutation on: (i) the binding of the 2-aminopyrimidine scaffold of osimertinib to EGFR hinge region within the ATP binding pocket, and (ii) the accommodation of the acrylamide warhead.

Analysis of the trajectories indicated that L718Q mutation has only a small effect on the overall stability of the EGFR protein. The root-mean square deviation (RMSD) of the backbone atoms for the first set of 300 ns of MD simulation was of $1.60 \pm 0.49 \text{ \AA}$ for EGFR T790M and $2.20 \pm 0.31 \text{ \AA}$ for EGFR T790M/L718Q. On the other hand, a significant increase in the flexibility of the P-loop, *i.e.* the glycine-rich stretch where the mutation is located, was observed when Leu718 was replaced by Gln. Backbone-atom RMSD values were $1.99 \pm 0.49 \text{ \AA}$ and $3.50 \pm 0.85 \text{ \AA}$ for EGFR T790M and EGFR T790M/L718Q, respectively. However, this increased flexibility of the P-loop did not affect the overall stability of the EGFR- osimertinib complex: the RMSD for the heterocyclic core of osimertinib was $1.13 \pm 0.32 \text{ \AA}$ for EGFR T790M and $1.62 \pm 0.49 \text{ \AA}$ for the EGFR T790M/L718Q double mutant. Also, the distance of the key H-bond formed by the N3-pyrimidine nitrogen of osimertinib and the backbone N-H group of Met793 remains essentially the same *i.e.*, $2.19 \pm 0.21 \text{ \AA}$ for EGFR T790M and $2.11 \pm 0.15 \text{ \AA}$ for EGFR T790M/L718Q double mutant. Similar behavior was observed in all the simulations (see Fig. S3–S6, ESI†).

Further analysis of the MD simulations showed that the presence of Gln instead of Leu did not dramatically alter the average distance between Cys797 sulfur atom and the C β of the acrylamide, which moves from $4.9 \pm 0.90 \text{ \AA}$ for EGFR T790M to $6.0 \pm 1.0 \text{ \AA}$ for EGFR T790M/L718Q. However, detailed analysis of the MD trajectories revealed that for EGFR T790M, the fraction of reactive conformations, *i.e.* configurations in which the

nucleophile (Cys797 S) and electrophile (acrylamide C β) are separated by a distance smaller than the sum of their van der Waals radii⁴² (*i.e.* 3.9 \AA with the AMBER99SB force field), corresponds to 20.5% of structures, while in the case of the EGFR T790M/L718Q double mutant, only 0.7% of structures have Cys797 sulfur and acrylamide C β at distances shorter than 3.9 \AA . Similar results were obtained in other MD replicas (Table S4†). In these trajectories, the distribution of S-C β distances depends on the values assumed by the rotatable bond connecting the nitrogen of the acrylamide to the phenyl ring of osimertinib (Fig. 4A). Analysis of the time series of S-C β distance and C₁-C₂-N₁-C₃ dihedral angle (which describes the rotation along the above-mentioned C-N bond) indicates that these two geometrical descriptors are highly correlated (Fig. 4). When the S-C β distance approached the critical value of 3.9 \AA , the dihedral angle assumed values of $-55^\circ \pm 20^\circ$, consistent with the value observed in the X-ray structure, while when S-C β distance overcame the value of 5.5 \AA , a nearly 180° shift in the C₁-C₂-N₁-C₃ dihedral angle was observed.

This trend was observed both in EGFR T790M (Fig. 4B) and in EGFR T790M/L718Q (Fig. 4C). The time series reported in Fig. 4B and C show the existence of two conformational states for EGFR- osimertinib non-covalent complexes. Similar trends

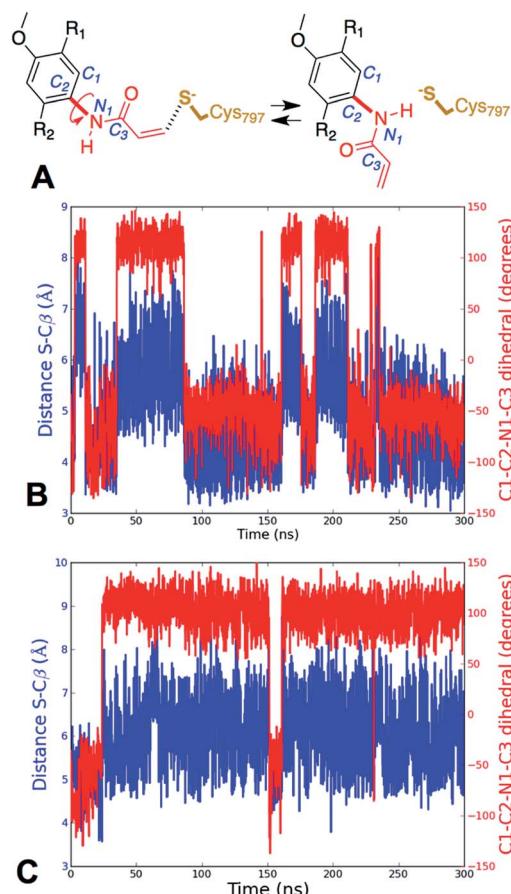


Fig. 4 Definition of C₁-C₂-N₁-C₃ dihedral angle (panel A), and time series for the S-C β distance (blue line) and the C₁-C₂-N₁-C₃ dihedral (red line) for EGFR T790M (panel B) and EGFR T790M/L718Q (panel C).



were also observed when the MD simulations were replicated (Fig. S7†).

Using MD simulations of the T790M and T790M/L718Q systems, we calculated conformational FESs along S-C β distance and the C₁-C₂-N₁-C₃ dihedral angle. These FESs were built by pooling the snapshots collected from four MD replicas each of 300 ns (*i.e.* from a total of 1.2 μ s MD simulations for each system) and then performing a histogram analysis of the conformational frequencies (see Experimental section). The resulting FESs confirmed the existence of two basins (labelled as **a** and **b**) for both EGFR T790M (Fig. 5A) and EGFR T790M/L718Q (Fig. 5B) mutants.

In the case of EGFR T790M, the region **a** identifies a well-defined minimum centered at a S-C β distance of 3.5 \AA and a C₁-C₂-N₁-C₃ dihedral angle of -55° . This corresponds to the geometry of the EGFR T790M–osimertinib complex belonging to basin **R** of Fig. 2. Region **a** thus contains reactive conformations, *i.e.*, conformations extending from the ground state of the complex that lie on the transition path leading to the alkylation of Cys797. Region **b** identified a wider minimum corresponding to unreactive conformations of the EGFR–osimertinib complex, in which the S-C β distance was dispersed over a 4.5–6.5 \AA interval due to a change in C₁-C₂-N₁-C₃ dihedral angle, which adopted conformations in the range $+100^\circ$ to $+120^\circ$. **a** and **b** possess nearly the same free-energy and thus they can be regarded as equally populated conformational states.

For EGFR T790M/L718Q, significant changes were observed in the positions of minima **a** and **b**, and in their relative populations. Firstly, minimum **a** does not correspond to any reactive conformations. Configurations in this area of the FES feature a S-C β distance approaching 4.5 \AA and thus are hardly capable of leading to Cys797 alkylation. Minimum **b** describes unreactive conformations where the S-C β distance is even higher (5.5–7.0 \AA range) while the C₁-C₂-N₁-C₃ dihedral angle remains close to the interval ranging from $+90^\circ$ to $+110^\circ$, already

observed for the unreactive basin of the T790M mutant. Importantly, basin **b** represents the global minimum of the computed FES for the EGFR T790M/L718Q system, being more stable than basin **a** by nearly 1 kcal mol $^{-1}$. It is still possible to identify reactive geometries for the EGFR T790M/L718Q mutant. These conformations are situated on a region of the FES, labelled with **a*** (Fig. 5B), which does not correspond to a free-energy minimum. **a*** is higher in energy than basin **b** by \approx 3 kcal mol $^{-1}$. The stabilization of basin **b** compared to **a** and **a*** in the L718Q mutant arises from a H-bond between the side chain of Gln718 and the carbonyl oxygen of the acrylamide group of osimertinib (Fig. 6). The hydrogen bond is maintained for a significant part of the MD simulations. Comparison of the time evolution for the H-bond distance, the C₁-C₂-N₁-C₃ dihedral angle (Fig. S8†) and the S-C β distance (Fig. S9†),

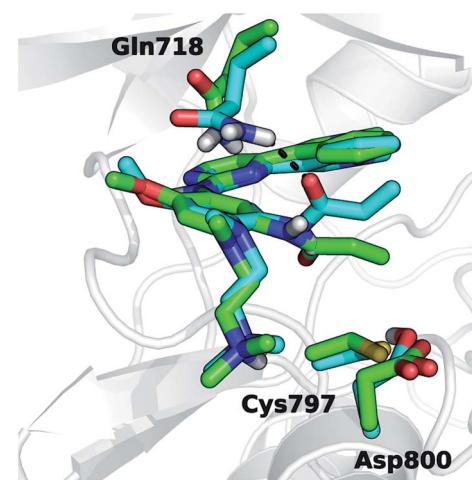


Fig. 6 Representative geometry structures for EGFR T790M/L718Q–osimertinib complexes taken from basin **a*** (green carbon atoms) and **b** (cyan carbon atoms) of Fig. 5B. The H-bond involving Gln718 and the acrylamide carbonyl is depicted with a dashed black line.

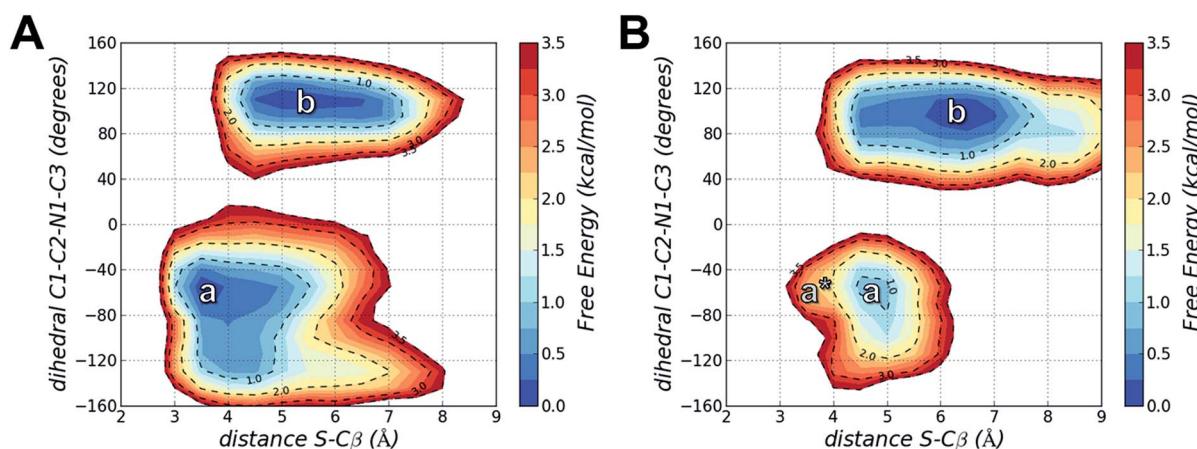


Fig. 5 Free energy surfaces calculated from the frequency distribution of conformations obtained from four independent 300 ns MD simulations of EGFR T790M (panel A) and EGFR T790M/L718Q (panel B) in non-covalent complex with osimertinib. **a** and **b** identify two conformational basins in the space of S-C β distance and C₁-C₂-N₁-C₃ dihedral angle; **a*** represents, in panel B, the region of reactive conformations for EGFR T790M/L718Q–osimertinib (which approximately corresponds to basin **a** in A). Free energies are given in kcal mol $^{-1}$, the contour levels are set at 0.25 kcal mol $^{-1}$ while dashed-contour lines are set every 1 kcal mol $^{-1}$ with an additional line at 3.5 kcal mol $^{-1}$.



indicates that formation of this H-bond is the key driver leading to stabilization of the unreactive states for EGFR T790M/L718Q–osimertinib complex. The hydrogen bond between the side-chain of Gln718 and the carbonyl oxygen of acrylamide drives the transition of the C₁–C₂–N₁–C₃ dihedral angle and places the electrophile C β far away from the Cys797 thiolate. This H-bond mediated stabilization of a non-reactive conformation; this conformational change of the non-covalent complex will hamper the formation of a covalent bond and thus appears to be the central cause of how the T790M/L718Q mutant escapes irreversible inhibition by osimertinib. These conclusions are supported by all the other MD replicas (which are reported in Fig. S8 and S9 of the ESI[†]).

Conclusions

The clinical use of third-generation EGFR inhibitors is revealing the insurgence of novel mutations which confer to lung cancer cells the ability to escape EGFR inhibition. Among recently reported mutations, L718Q has emerged as an intriguing mutation considering that this occurs in a region of EGFR peripheral to the ATP binding site and that its effect is not obvious (in contrast *e.g.* to the C797S mutation). Here, a multilevel computational approach has provided evidence that L718Q mutation does not affect the ionization state of Cys797 (the intrinsic reactivity of Cys797 is not reduced) nor the activation free-energy for the direct addition of acrylamide to the α,β double bond. Free-energy calculations, performed with a replica exchange-thermodynamic integration method, indicate that L718Q mutation does not affect the affinity of osimertinib for the EGFR active site. On the other hand, simulations show that Gln718, through an H-bond with the acrylamide warhead, stabilizes a specific conformation of osimertinib in which the electrophilic C β of its acrylamide substituent is kept away from the Cys797 thiolate. This conformation hampers Cys797 alkylation and may thus favor the displacement of osimertinib from EGFR by ATP, which is present in cells at high concentrations.¹¹ This global picture rationalizes the poor inhibitory activity of osimertinib in Ba/F3 cells harboring the L718Q mutation¹⁸ and the lack of therapeutic activity of osimertinib in patients who have acquired this mutation.¹⁹ We believe that the protocol described here represents a viable and generalizable strategy to investigate *in silico* the effects of mutations on covalent inhibition. This requires consideration of several distinct factors: (i) evaluation of the ionization state of the nucleophile; (ii) estimation of the free-energy barrier for covalent adduct formation; (iii) binding affinity for the formation of the non-covalent complex; (iv) mutual orientation of the reactants within the target active site. Moreover, our conclusion about the conformational selection induced by L718Q mutation should be useful for the design of new EGFR inhibitors overcoming drug resistance.

Experimental

Model building and equilibration

EGFR T790M and EGFR T790M/L718Q in complex with osimertinib were built starting from the crystal structure of the

non-covalent complex between wt EGFR and osimertinib (4ZAU.pdb),²⁰ using the LEaP program available in AMBER 16 (AMBER 2016, University of California, San Francisco, CA). The resulting complexes were immersed in a box of TIP3P water molecules⁴³ and neutralized with 3 chloride ions. The total system size was 46 936 atoms (14122 TIP3P waters) in the case of the EGFR T790M–osimertinib complex and 46 283 atoms in the case of the EGFR T790M/L718Q–osimertinib complex (13905 waters). Solvated complexes were energy-minimized, gradually heated to 300 K in NVT ensemble and equilibrated at pressure of 1 atm in NPT ensemble. These systems were then submitted to a short MD simulation (10 ns) in the NVT ensemble. In all these simulations, the AMBER99SB force field was applied to describe protein atoms while the generalized AMBER force field (GAFF)⁴⁴ was employed to describe osimertinib. The pmemd.cuda code of AMBER16 was used to perform these simulations on NVIDIA K20 GPU cards. Full electrostatic and van der Waals interactions were computed within a cutoff of 10 Å and long range electrostatic interactions were treated using the particle mesh Ewald (PME) with 128 \times 128 \times 128 grid points. Covalent bonds involving hydrogens were constrained using the SHAKE function, allowing the use of a time step of 2 fs. The structures extracted from these MD trajectories (snapshots) were employed to estimate pK_a for Cys797, for QM/MM simulations, and to perform further MD simulations (see below).

pK_a calculations

Ten equally spaced snapshots were taken from the 10 ns MD trajectories for EGFR T790M–osimertinib and EGFR T790M/L718Q–osimertinib complexes. The structures were exported as PDB files and submitted to pK_a calculations using Propka3.0 (ref. 32) and H++ (ref. 33) software using default settings. Individual ΔpK_a values were calculated as Cys797 pK_a of EGFR T790M minus Cys797 pK_a of EGFR T790M/L718Q, while the final values reported in main text resulted from their average. The uncertainty of ΔpK_a was estimated calculating the standard error of the mean (SEM). Absolute pK_a values are reported in Table S1 in the ESI.[†]

Application of the QM/MM potential

QM/MM has been extensively applied to study reaction mechanisms in condensed phase, including enzymes. This approach has the advantage that large systems can be investigated with a reasonable computational cost. The central idea behind the QM/MM approach is the partitioning of the system of interest into two interacting regions: (i) a small QM region, where the chemical bond breaking and forming take place and (ii) a large MM region surrounding the QM atoms. Here, we applied the self-consistent charge-density functional tight binding (SCC-DFTB)⁴⁵ model to describe the QM region and the AMBER99SB force field to describe the MM region. The SCC-DFTB approach is based on the second-order expansion of the total DFT energy with respect to the charge-density variation. Hybrid QM/MM potentials derived from the SCC-DFTB theory have been widely applied to enzyme-catalyzed reactions and often give



reasonable descriptions of reaction geometries and energetics.⁴⁶ In the EGFR–osimertinib complexes, side chain atoms of Cys797, Asp800 and the acrylamide portion of the inhibitor were treated with the SCC-DFTB method with dispersion correction. The resulting QM system was composed of 23 atoms, including three link atoms placed along the C–C bond connecting C β of the amino acids to their backbone C α and along the C–N bond connecting the acrylamide nitrogen to the phenyl ring of osimertinib. The `adjust_q` function was applied to conserve the charge of the systems.

US simulation for Cys797 deprotonation by Asp800

A set of US simulations was performed by adding a spring constant of $k = 100$ kcal mol $^{-1}$ Å $^{-2}$ along the difference of distances (H_{Asp800}–S_{Cys797}) – (H_{Asp800}–O_{Asp800}). This difference was named as reaction coordinate r and was sampled from 1 to –1 Å with a step-size of 0.1 Å, giving a total of 21 windows. For each window, 50 ps of QM/MM MD simulation at 300 K in NVT conditions were performed for a total of 1 ns of simulation, using the `sander.MPI` module of AMBER16. All the atoms of the system (including hydrogens) were allowed to move during the simulations, and a time step of 0.2 fs was used to integrate the equation of motion. The PME approach was used to treat the QM/MM long-range electrostatic interactions. The unbiased FES was calculated by using the weighted histogram analysis method (WHAM).⁴⁷ Convergence of the simulations was evaluated calculating the difference in the free-energy between Cys797–S $^-$ /Asp800–COOH and Cys797–SH/Asp800–COO $^-$ states as function of the simulation time. This difference reached a stable value after 40 ps of simulation for each window (Fig. S10†). The free-energy difference between the two ionization states was calculated at different simulation time once convergence was achieved *i.e.*, after 40, 45 and 50 ps. The final energies reported here represent mean values \pm standard deviation.

US simulation for Cys797 alkylation

US MD simulations were carried out adding a spring constant of 100 kcal mol $^{-1}$ Å $^{-2}$ along the reaction coordinate x , described by the S_{Cys797}–C β acrylamide distance, and reaction coordinate y , described by the difference of distances (H_{Asp800}–C α acrylamide) – (H_{Asp800}–O_{Asp800}). Coordinate x was sampled from 3.6 to 1.8 Å with a step-size of 0.1 Å for a total of 19 windows while coordinate y was sampled from 2.8 to –2.0 Å with a step-size of 0.2 Å for a total of 26 windows. Overall the simulated grid is composed of 494 windows. For each one, 50 ps of QM/MM MD simulation at 300 K in NVT conditions were performed for a total of 24.7 ns. The same settings described for Cys797–Asp800 proton transfer were applied. The presence of a normal distribution of the sampled data along x and y was verified within each simulated window. The employed spring constant allowed a satisfactory overlapping of the sampled windows, which is critical to obtain a reliable and continuous FES performed for each umbrella. The unbiased FES was calculated by using the weighted histogram analysis method (WHAM).⁴⁷ Convergence of the simulations was evaluated calculating the

ΔA_{act} and ΔA_{reac} as functions of the simulations time. These two outputs reached stable values within 30 ps of QM/MM MD simulation for each window. Average values and standard deviations reported in the main text were calculated from data collected at 30, 40 and 50 ps of simulation for each window. This was the case of both EGFR–osimertinib complexes (Fig. S11†). The US simulations were also repeated to evaluate their reproducibility (Fig. S12†).

WaterSwap calculations

From a 300 ns trajectory for the EGFR T790M–osimertinib and EGFR T790M/L718Q–osimertinib complexes, 10 equally spaced snapshots were selected and employed for WaterSwap calculations. Binding free energies were calculated using replica-exchange thermodynamic integration over 16λ windows (0.005, 0.071, 0.137, 0.203, 0.269, 0.335, 0.401, 0.467, 0.533, 0.599, 0.665, 0.731, 0.797, 0.863, 0.929, 0.995) over the WaterSwap reaction coordinate. 30 million Monte Carlo moves were performed for each window, with the free-energy gradient averaged over the last 20 million steps. Monte Carlo sampling involved protein residues and water molecules within 15 Å of the center of osimertinib. Simulations used the set A soft-core parameters described in ref. 39, with a 15 Å coulomb and Lennard Jones non-bonded cutoff. This shifted cutoff was used to better account for long-range electrostatics.

ΔA_{bind} values reported here are mean values \pm the standard error of the mean, as obtained from 10 independent simulation runs.

Unbiased MD simulations and free energy calculations

Starting from equilibrated systems of EGFR T790M–osimertinib and EGFR T790M/L718Q–osimertinib complexes we performed 4 runs of 300 ns-long MD simulations at 300 K in NVT conditions using the settings described above. Snapshots collected from MD trajectories (30 000 for each replica, for a total of 120 000 for each system) were employed to build the 2D-free-energy surfaces in the space of S–C β distance and C₁–C₂–N₁–C₃ dihedral angle. The free-energy along these variables was computed applying the following equation:

$$A(r) = -k_B T \ln P(r)$$

where P is the probability distribution, k_B is the Boltzmann constant, and T is the simulation temperature. The 2D distribution function $P(r)$ was obtained performing a histogram analysis of conformational frequencies with the `cpptraj` module⁴⁸ implemented in AMBER16. Snapshots were separated into classes based on the values of the S–C β distance and C₁–C₂–N₁–C₃ dihedral angle. S–C β distance ranged from 2.5 to 9.5 Å with a class width of 0.1 Å, while the C₁–C₂–N₁–C₃ dihedral angle ranged from –160° to +160° with a class width of 15°. FESs in the space of S–C β distance and C₁–C₂–N₁–C₃ dihedral angle for each replica are reported for EGFR T790M–osimertinib and EGFR T790M/L718Q–osimertinib complexes in Fig. S13 and S14, respectively of the ESI† section.



Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was carried out using the computational facilities of the Advanced Computing Research Centre, University of Bristol – <http://www.bris.ac.uk/acrc/>. This research also benefits from the HPC (High Performance Computing) facility of the University of Parma, Italy - <http://www.hpc.unipr.it>. This work is supported by CCP-BioSim (www.cepbiosim.ac.uk; EPSRC grant number EP/M022609/1 to AJM) and by BBSRC grant number BB/M000354/1 (to AJM and KER). PhD scholarship of DC is provided by the University of Parma. CJW is an EPSRC Research Software Engineering Fellow (EP/N018591/1). CJW acknowledges the support of BrisSynBio, a BBSRC/EPSRC Synthetic Biology Research Centre (BB/L01386X/1).

Notes and references

- 1 N. E. Hynes and H. A. Lane, *Nat. Rev. Cancer*, 2005, **5**, 341–354.
- 2 J. Schlessinger, *Cell*, 2000, **103**, 211.
- 3 N. Normanno, A. De Luca, C. Bianco, L. Strizzi, M. Mancino, M. R. Maiello, A. Carotenuto, G. De Feo, F. Caponigro and D. S. Salomon, *Gene*, 2006, **366**, 2.
- 4 S. V. Sharma, D. W. Bell, J. Settleman and D. A. Haber, *Nat. Rev. Cancer*, 2007, **7**, 169.
- 5 W. Pao and J. Chmielecki, *Nat. Rev. Cancer*, 2010, **10**, 760.
- 6 C. Carmi, A. Lodola, S. Rivara, F. Vacondio, A. Cavazzoni, R. Alfieri, A. Ardizzone, P. G. Petronini and M. Mor, *Minirev. Med. Chem.*, 2011, **11**, 1019.
- 7 J. A. Engelman and P. A. Jänne, *Clin. Cancer Res.*, 2008, **14**, 2895.
- 8 A. Michalczyk, S. Klüter, H. B. Rode, J. R. Simard, C. Grüter, M. Rabiller and D. Rauh, *Bioorg. Med. Chem.*, 2008, **16**, 3482.
- 9 C. H. Yun, K. E. Mengwasser, A. V. Toms, M. S. Woo, H. Greulich, K. K. Wong, M. Meyerson and M. J. Eck, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 2070.
- 10 V. Hirsh, *Future Oncol.*, 2011, **7**, 817–825.
- 11 C. Carmi, M. Mor, P. G. Petronini and R. Alfieri, *Biochem. Pharmacol.*, 2012, **84**, 1388.
- 12 N. Katakami, S. Atagi, K. Goto, T. Hida, T. Horai, A. Inoue, Y. Ichinose, K. Koboyashi, K. Takeda, K. Kiura, K. Nishio, Y. Seki, R. Ebisawa, M. Shahidi and N. Yamamoto, *J. Clin. Oncol.*, 2013, **31**, 3335.
- 13 M. R. V. Finlay, M. Anderton, S. Ashton, P. Ballard, P. A. Bethel, M. R. Box, R. H. Bradbury, S. J. Brown, S. Butterworth, A. Campbell, C. Chorley, N. Colclough, D. A. E. Cross, G. S. Currie, M. Grist, L. Hassall, G. B. Hill, D. James, M. James, P. Kemmitt, T. Klinowska, G. Lamont, S. G. Lamont, N. Martin, H. L. McFarland, M. J. Mellor, J. P. Orme, D. Perkins, P. Perkins, G. Richmond, P. Smith, R. A. Ward, M. J. Waring, D. Whittaker, S. Wells and G. L. Wrigley, *J. Med. Chem.*, 2014, **57**, 8249.
- 14 S. S. Ramalingam, J. C. Yang, C. K. Lee, T. Kurata, D. W. Kim, T. John, N. Nogami, Y. Ohe, H. Mann, Y. Rukazekov, S. Ghiorgiu, D. Stetson, A. Markovets, J. C. Barrett, K. S. Thress and P. A. Jänne, *J. Clin. Oncol.*, 2017, DOI: 10.1200/JCO.2017.74.7576.
- 15 D. A. Cross, S. E. Ashton, S. Ghiorgiu, C. Eberlein, C. A. Nebhan, P. J. Spitzler, J. P. Orme, M. R. Finlay, R. A. Ward, M. J. Mellor, G. Hughes, A. Rahi, V. N. Jacobs, M. Red Brewer, E. Ichihara, J. Sun, H. Jin, P. Ballard, K. Al-Kadhim, R. Rowlinson, T. Klinowska, G. H. Richmond, M. Cantarini, D. W. Kim, M. R. Ranson and W. Pao, *Cancer Discovery*, 2014, **4**, 1046.
- 16 J. Engel, J. Lategahn and D. Rauh, *ACS Med. Chem. Lett.*, 2016, **7**, 2.
- 17 K. S. Thress, C. P. Paweletz, E. Felip, B. C. Cho, D. Stetson, B. Dougherty, Z. Lai, A. Markovets, A. Vivancos, Y. Kuang, D. Ercan, S. E. Matthews, M. Cantarini, J. C. Barrett, P. A. Jänne and G. R. Oxnard, *Nat. Med.*, 2015, **21**, 560.
- 18 D. Ercan, H. G. Choi, C. H. Yun, M. Capelletti, T. Xie, M. J. Eck, N. S. Gray and P. A. Jänne, *Clin. Cancer Res.*, 2015, **21**, 3913.
- 19 M. Bersanelli, R. Minari, P. Bordi, L. Gnetti, C. Bozzetti, A. Squadrilli, C. A. Lagrasta, L. Bottarelli, G. Osipova, E. Capelletti, M. Mor and M. Tiseo, *J. Thorac. Oncol.*, 2016, **11**, e121.
- 20 Y. Yosaatmadja, S. Silva, J. M. Dickson, A. V. Patterson, J. B. Smaill, J. U. Flanagan, M. J. McKeage and C. J. Squire, *J. Struct. Biol.*, 2015, **192**, 539.
- 21 (a) L. Sutto and F. L. Gervasio, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 10616; (b) J. Park, J. J. McDonald, R. C. Petter and K. N. Houk, *J. Chem. Theory Comput.*, 2016, **12**, 2066.
- 22 M. De Vivo, M. Masetti, G. Bottegoni and A. Cavalli, *J. Med. Chem.*, 2016, **59**, 4035.
- 23 W. Sinko, S. Lindert and J. A. McCammon, *Chem. Biol. Drug Des.*, 2013, **81**, 41.
- 24 (a) M. J. Field, P. A. Bash and M. Karplus, *J. Comput. Chem.*, 1990, **11**, 700; (b) A. Warshel, *Annu. Rev. Biophys. Biomol. Struct.*, 2005, **32**, 425; (c) A. Lodola and M. De Vivo, *Adv. Protein Chem. Struct. Biol.*, 2012, **87**, 337; (d) A. Lodola and A. J. Mulholland, *Methods Mol. Biol.*, 2013, **924**, 67.
- 25 C. J. Woods, M. A. King and J. W. Essex, *J. Phys. Chem. B*, 2003, **107**, 13703.
- 26 (a) A. Berteotti, F. Vacondio, A. Lodola, M. Bassi, C. Silva, M. Mor and A. Cavalli, *ACS Med. Chem. Lett.*, 2014, **5**, 501; (b) M. E. Flanagan, J. A. Abramite, D. P. Anderson, A. Aulabaugh, U. P. Dahal, A. M. Gilbert, C. Li, J. Montgomery, S. R. Oppenheimer, T. Ryder, B. P. Schuff, D. P. Uccello, G. S. Walker, Y. Wu, M. F. Brown, J. M. Chen, M. M. Hayward, M. C. Noe, R. S. Obach, L. Philippe, V. Shanmugam, M. J. Shapiro, J. Starr, J. Stroh and Y. Che, *J. Med. Chem.*, 2014, **57**, 10072.
- 27 P. A. Schwartz, P. Kuzmic, J. Solowiej, S. Bergqvist, B. Bolanos, C. Almaden, A. Nagata, K. Ryan, J. Feng, D. Dalvie, J. C. Kath, M. Xu, R. Wani and B. W. Murray, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 173.
- 28 P. Madzelan, T. Labunská and M. A. Wilson, *FEBS J.*, 2012, **279**, 4111.



29 T. H. Truong and K. S. Carroll, *Biochemistry*, 2012, **51**, 9954.

30 L. Capoferri, A. Lodola, S. Rivara and M. Mor, *J. Chem. Inf. Model.*, 2015, **55**, 589.

31 T. H. Truong, P. M. Ung, P. B. Palde, C. E. Paulsen, A. Schlessinger and K. S. Carroll, *Cell Chem. Biol.*, 2016, **23**, 837.

32 M. H. Olsson, C. R. Søndergaard, M. Rostkowski and J. H. Jensen, *J. Chem. Theory Comput.*, 2011, **7**, 525.

33 R. Anandakrishnan, B. Aguilar and A. V. Onufriev, *Nucleic Acids Res.*, 2012, **40**, W537.

34 G. M. Seabra, R. C. Walker, M. Elstner, D. A. Case and A. E. Roitberg, *J. Phys. Chem. A*, 2007, **111**, 5655.

35 A. Paasche, M. Schiller, T. Schirmeister and B. Engels, *ChemMedChem*, 2010, **5**, 869.

36 (a) D. G. Xu and H. Guo, *J. Am. Chem. Soc.*, 2009, **131**, 9780; (b) A. Lodola, D. Branduardi, L. Capoferri, M. De Vivo, M. Mor and D. Piomelli, *PLoS One*, 2012, **7**, e32397; (c) L. L. Pan, L. F. Song, Y. Miao, Y. Yang and K. M. Merz, *Biochemistry*, 2017, **56**, 2995.

37 W. P. Jencks, *Chem. Rev.*, 1972, **72**, 705.

38 X. E. Yan, S. J. Zhu, L. Liang, P. Zhao, H. G. Choi and C. H. Yun, *Oncotarget*, 2017, **8**, 53508.

39 C. J. Woods, M. Malaisree, S. Hannongbua and A. J. Mulholland, *J. Chem. Phys.*, 2011, **134**, 054114.

40 V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, *Proteins*, 2006, **65**, 712.

41 C. J. Woods, M. Malaisree, B. Long, S. McIntosh-Smith and A. J. Mulholland, *Sci. Rep.*, 2013, **3**, 3561.

42 R. Lonsdale, K. E. Ranaghan and A. J. Mulholland, *Chem. Commun.*, 2010, **46**, 2354.

43 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L Klein, *J. Chem. Phys.*, 1983, **79**, 926.

44 J. M. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157.

45 M. Elstner, *Theor. Chem. Acc.*, 2006, **116**, 316.

46 M. Gruden, L. Andjeklović, A. K. Jissy, S. Stepanović, M. Zlatar, Q. Cui and M. Elstner, *J. Comput. Chem.*, 2017, **38**, 2171.

47 S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen and P. A. Kollman, *J. Comput. Chem.*, 1992, **13**, 1011.

48 D. R. Roe and T. E. Cheatham, *J. Chem. Theory Comput.*, 2013, **9**, 3084.